Global-scale analysis for river flow alterations due to water withdrawals and reservoirs

P. Döll, K. Fiedler, and J. Zhang

Final response to referee comments

We thank Balazs Fekete and two anonymous referees for their very good comments which helped us to improve our manuscript. In the following, we answer to the comments of the reviewers.

Referee 1 (Fekete)

RC1: Döll et al. presented a new study based on the WaterGAP Global Hydrological Model (WGHM) simulation to assess to provide a thorough assessment of the anthropological flow alterations global. WGHM is undoubtedly a state-of-the art model to carry out the presented research. The authors run the model in several configuration turning on and off various human disturbances (reservoir operation, irrigational water uptake).
AC1: not only irrigation water use but total water use was "turned on and off"

RC2: It is somewhat unclear, how the model was calibrated against the long term mean flow conditions at 1235 station that likely experienced varying degree of human flow alteration during their operation.
AC2: The following text was added to section 2.1.1. "During tuning, observed long-term average discharge at the basin outlet was compared to modeled long-term average discharge during the observation time periods. Modeled discharge took into account the time-varying consumptive use during the tuning period, but it was assumed that the reservoirs existed during the whole tuning period."

RC3: The presented work is still valuable since it gives a detailed account of how flow characteristic change due to human activities. The tested flow indicators go well beyond the ones normally used in global scale studies. To some degree, one has to question if WGHM is indeed capable to reproduce these characteristics realistically. The authors spent a great deal on showing the model performance in terms of reservoir operation and the WGHM performance as a water balance model was documented in the past.
What is missing is the demonstration that the low flow frequencies, the flow variation amplitudes, flow regime, time shifts, inter-anual variability of the monthly flow (the indicators that the authors evaluated) are indeed reasonable approximation of the real world. Without having prior and post human alteration discharge records the demonstration of the model capabilities is obviously difficult, but the authors should show at least a comparison to observed time series data that the model actually captures these
characteristic realistically.
AC3: In Döll et al. (2003) we showed that low flow indicator Q90 is reasonable well simulated by WGHM, while modelled and observed time series of monthly river discharges are in some basins quite close and in others very different. In the course of the presented study, we also looked at the capability of WGHM to represent interannual variability. We found that for selected basins, this capability may differ strongly dependent on the precipitation data set used (GPCC, CRU 2.1 or CRU 3.0). And again, in some basins WGHM does a good job, while in others, it does a terrible job. To communicate this to the readers, we added the following sentences to section 2.2:
"Please note, however, that nevertheless the uncertainty of the computed indicators is very high. When comparing observed to simulated (ANT) values of the flow regime indicators used to derive the indicators of river flow alteration (Q_{90} as well as mean and interannual variability of monthly flows), the fit is good in some basins and bad in others, often depending on the applied global precipitation data set."

RC4: To some degree, I doubt if a monthly model can depict shifts in discharge records that are likely to be in the order of weeks rather than months. The tested model discharge gauges (Figure 9) clearly show that WGHM has limited ability to capture the flow amplitude and the flow regime. Since, the model runs at daily time step internally (if I understood correctly), one has to wonder, why the authors did not try to evaluate these metrics at the computed daily time step. I realize that the precipitation downscaling the authors used (based on wet days frequencies) is not sufficient to generate reasonable daily output, but combining the GPCC full product with daily precipitation from NCEP or ERA40 reanalysis (preserving the monthly totals) could provide realistic daily precipitation input.
AC4: In future work, we will evaluate river discharge computed from (almost) daily climate input but it is necessary to correct the number of wet day in reanalysis as these are overestimated.

RC5: The incorporation of the 6000+ GRanD reservoirs in this study is somewhat unclear.The GRanD reservoirs spread over a 30-minute network (roughly 65000 continental grid cells) would take up ten percent of the grid cells if they were distributed as one reservoir per grid cell. Obviously the authors performed some degree of lumping multiple reservoirs sharing the same grid cells. I suppose the 1074 reservoirs where the authors actually applied the reservoir operation (according to Hanasaki et al.) represents some sort of combined reservoirs operating per grid cell.
AC5: This is correct. To clarify this, the first paragraph of section 2.1 was reformulated as follows:
"For each grid cell of WGHM, a vertical water balance is computed, and the resulting runoff is routed laterally within the cell through a groundwater store and various surface water stores (if existent: lakes, reservoirs, wetlands, always assumed to exist: river,). The effect of surface water storage on water balance and flow dynamics is modeled by first routing the runoff generated within the grid cell through a so-called "local" lake/reservoir storage and a "local" wetland storage compartment. The resulting discharge volume is added to the discharge from the upstream grid cell and routed through a so-called "global" lake and/or reservoir compartment and a "global" wetland storage compartment, and finally through the river storage compartment. If there are a number of lakes, reservoirs or wetlands within each grid cells they are lumped into one. In the "local" case, lakes and reservoirs are lumped together, too, while in the "global" case, lakes and reservoirs are distinguished. The water balance of "global" lakes and reservoirs, which can cover more than one grid cell, is performed at the grid cell where the outflow of the lake or reservoir is located. The difference between precipitation and potential evapotranspiration is added to the water balance of lakes, reservoirs and wetland, thus taking into account the effect of the surface water balance on cell runoff. "

RC6: To some degree, one has to wonder, why the authors went through all the hoops to test so many metrics and at the and to use simple discharge vs. fish species relationship to assess the lost biodiveristy, due to human activities.
AC6: As noted in the last paragraph of the paper, the metrics beyond long-term average river discharge will have to be the basis for improved relations between biodiversity changes and river flow alterations (beyond the only currently available quantitative relation between a flow metric and a biodiversity metric, the fish species numbers vs. long-term average river discharge relation)

RC7: The reality is that over 6 billion people live on our planet that will reach 9 billion in our lifetime. With all my sympathy to endangered species, if I had to choose between feeding people or saving fish, I would undoubtedly choose the first. To some degree, these studies should try to inform us, how much the human alterations are "wasteful" and how much inevitable.
AC7: Arguably, enough food could be produced globally without any irrigation (according to calculations of Siebert and Döll, J.Hydrol., 2009, global cereal production would decrease by only on quarter without irrigation, and such decreases could be balanced by yield increases,

and virtual water trade). This, of course, neglects the realities of income generation necessary to buy food. Thus, it is e.g. a question of political will and societal values which human alterations are considered to be “inevitable”.

Referee 2
General comments
RC8: WGHM is a calibrated model. It is not clearly stated in the paper how the calibration was handled in the different model runs. Several clarifications are needed about the calibration. Is the model calibrated to the standard (ANT) run, or is any calibration made also for e.g. the naturalized flow run? If the latter is not the case, which I suppose, please discuss the possible impact of your results for the ANT LAKE, USE, RES and USE runs when the calibrated parameter value comes from the ANT case. And if not, please discuss the implication when comparing results from runs with different calibration parameters. Was the model calibrated again compared to the Hunger and Döll paper, and what were in that case the differences?
AC8: Calibration was made for the standard run (ANT) only, as observation data reflect the impact of withdrawals and reservoirs, and there are (almost no) no observation data for the time period in which basins were without reservoirs and water withdrawals. For this study, WGHM 2.1g was run with calibration parameters obtained from calibrating WGHM 2.1 because we expected that parameters would not be affected much by the additional reservoirs (which increased reservoir surface from 254000 km2 to 291000 km2) in particular because many of the former lakes in 2.1f were relabelled as reservoirs in 2.1g. Nevertheless, the calibration parameters differ from those used in Hunger and Döll (2008) as different precipitation data sets were used. Please compare AC2 to the comments of B. Fekete.

RC9: Were the four validation basins used in the calibration or not? If not, was there any calibration made downstream in those basins, or are the gauging stations totally independent?
AC9: The four validation basins are part of the 1235 calibration basins. Added in 4 Discussion.

RC10: The paper is well written, but personally do I prefer a somewhat different order of presentations. Answers to my question marks often come one or two sentences below. As an example the model efficiency is introduced 2-3 rows above the specification that this is the Nash-Sutcliffe coefficient.
AC10: In case of the example given, the order was modified.

Specific comments:

RC11: Abstract, l. 6: Stating that this is the first study is for me a little bit too strong, thinking of e.g. Vörösmarty et al. (1997). You may still use “first” but please rewrite such that your particular contribution (flow variability) comes in closer connection to “first”.
AC11: The sentence now reads “This study presents the first global assessment of the anthropogenic alteration of river flow regimes, in particular of flow variability, by water withdrawals and dams.”

RC12: 4777, l. 15: 4000 km3/yr, is this value consistent between different estimates (WaterGAP, the model by Hanasaki, purely data-based estimates etc.) or does it vary a lot?
AC12: There are certain variations but we prefer not to discuss this in the text. Consumptive use is the important variable in our study.

RC13 4778, l. 20: Is the version 2.1g introduced with this paper? Please clarify this, and that Hunger and Döll (2008) is using version 2.1.f. Does “For this study . . .” (l. 22) mean changes made to the model in version 2.1g? It is not sufficient that you write more about the version numbers on 4779 l. 26.

AC13: The text was modified as follows: We used the global hydrology and water use model WaterGAP (Alcamo et al., 2003b) which takes into account the impact of reservoirs and water withdrawals on river discharge. For this study, we applied the most recent model version 2.1g. It differs from the previous version 2.1f as presented by Hunger and Döll (2008) with respect to the implementation of the reservoir algorithm of Hanasaki et al. (2006), and of the new GRanD reservoir data set (Lehner et al., 2008; Lehner et al. 2009, in preparation). With this version, the impact of more than 6500 reservoirs and regulated lakes could be analyzed.

RC14: 4781, l. 8: 2002? But the discharge is calculated for 1961-1990. What effect does this discrepancy in time have? And, how come that you later in this section show results for 1951-2002? How were these values calculated, and why do you not use the data from 1961-1990 in connection to the simulated runoff? Please inform better about the data sources you have used and their coverage in time, and please also discuss what effects the miss-matches in time might have. Do you try to explain something about this in the beginning of page 4785? If so, please move it to earlier in the methods description (it is actually hard to remember that the paper still is in the methods section after the references to figure 1, which easily are interpreted as results, and not background information).
AC14: To motivate the choice of time, we added to the second but last paragraph of the 1 Introduction: "We wanted to represent the alterations to natural conditions that had occurred by around the year 2000 due to withdrawals and dams only, i.e. under climatic conditions that have not yet been appreciably altered by climate change (i.e. before 1990)." Temporal development of consumptive water use is shown in an illustrative manner, to show how 2002 water use shown as a map in Fig. 1a relates to water use in former times. Besides, the time series of consumptive use between 1951 and 2002 is used for model tuning. Altogether, there is not really any temporal mismatch in the theoretical analysis presented in the paper, except when we compare our modelling results ANT, obtained with the water use of 2002, to observed reservoir outflows (as discussed on the top of page 4785).

RC15: Additionally, your statement on p. 4781, l. 25-26 would be easier to understand if Fig 1b. also include information on dry years (maybe as thin, vertical lines), but then you also have to define what you mean by "dry years".
AC15: Reference to Fig. 1b has been deleted as dry years cannot really be defined globally and are thus not well identifiable in Fig. 1b.

RC16: 4782, l. 10-11: Are there cells which are emptied and stay emptied until the end of the simulation?
AC16: Yes, e.g. in Saudi Arabia, with fossil gw use.

RC17: 4783, l. 6: What is the reason for not checking the smaller reservoirs? Too time consuming, data not accessible or are they really un-important? How do you expect the result to be influenced if these also had been included?
AC17: Too time consuming. Probably slightly increased alterations would have been computed. We also added the following to 4 Discussion: "The actual number of reservoirs, in particular the number of small reservoirs, is much higher than the number of reservoirs represented in WGHM. Therefore, the impact of reservoirs and regulated lakes on river flow regimes is certainly underestimated in this study."

RC18: 4783, l. 23: Why did you use different algorithms for the global and local reservoirs? Please explain.
AC18: lumping of many small reservoirs that might exist in one 0.5 grid cell into 1 effective local reservoirs makes reservoir modelling difficult (one reservoir is an irrigation reservoir, the other is a non-irrigation reservoir) and less beneficial than modelling the large global reservoirs as reservoirs. Very rarely there is more than one global reservoir per cell. Besides, definition e.g. of the operational year is time consuming. We added the sentence: "One reason for the latter is that the required lumping of various local reservoir within a grid cell

into one local reservoir per cell necessarily leads to a “blurring” of the specific reservoir characteristics, such that for local reservoirs, the reservoir algorithm is not expected to lead to more realistic results than the lake algorithm.”

RC19: 4784, l. 6: “five downstream cells”. Did Hanasaki et al. also use 0.5 degree cells, or did they use the larger 1 degree cells, i.e. does this procedure affect about the same area? It is somewhat unclear in this section what is different in the methods of Hanasaki et al. and your study and what is common.
AC19: The paragraph on the reservoir algorithm was expanded. See AC36

RC20: 4786, l. 8: “in river” => in large river. If they did not study large river basins, are the results actually applicable here?
AC20: no changes made. Results are applicable, as they included all sizes of basins and we estimate fish species values for the differently sized upstream areas of grid cells.

RC21: 4786, l. 22: Is this equation applicable for the upstream basin of each cell, when it was calculated for the basin outlet only? Additionally, were there any unregulated rivers in the work by Xenopoulos et al.? If not, is the equation applicable when using naturalized flows?
AC21: Yes, see AC20. Xenopoulos et al. assume that their equation reflects the evolutionary and ecological outcomes roughly in equilibrium with natural discharge.

RC22: 4787, l. 8: Do USE and RES sum up to the NAT values? Why, or why not?
AC22: Yes, approximately to the first decimal number, 0.8% + 0.2.7% = 3.5%, and non-linearities in individual basins seem to average out globally.

RC23: 4793, l. 13: Are the effects non-additive for individual basins, but the results even out on the global scale?
AC23: Yes, see AC22.

RC24: 4787, l. 18: Is this seen in measurements of the runoff too?
AC24: I don’t know.

RC25: 4788, l. 3: Is this decrease simulated by the model, or taken from some reference?
AC25: Not simulated from the model but a generally known and observed effect e.g. along the Tarim in Northwest China.

RC26: 4790, l. 5: Seems to be contradictory to the delays of several months reported by Vörösmarty et al. 1997? Please explain. Additionally, how did you calculate the area on row 8? Is it only the area of the river, downstream of the dam as these values are so low?
AC26: Vörösmarty computed increases of residence times (in river and reservoirs storage boxes that are assumed to be well-mixed containers) which is different from temporal shifts in flow peaks. The area of the grid cells downstream of the reservoir. In the third paragraph of 3 Results, we have added in the revised version: “With “16% of the global land” area we mean that 0.5° grid cells which cover 16% of the global land are are affected by this decrease. “

RC27: 4791, l. 20: Estimate too low also where you calculated a reduction of 99%?
AC27: I do not understand the question.

RC28: 4792, l. 12: Is the dam building/operation start time included in the model, such that the regulation is not used before this date?
AC28: No, and also not necessary for the type of analysis done in this study which compares two conditions: with and without dams.

RC29: 4795, l. 7: What is different with ITS? (Should probably be discussed in the discussion section and not here.)
AC29: I cannot find ITS here, or understand the question.

RC30: 4795: Acknowledgement. Are the data by e.g. Haddeland freely available over the web or have you received them personally? In the latter case I think it is appropriate to add it in the acknowledgement, or otherwise you should provide a link. If any of your data providers requires to be mentioned in the acknowledgement, all should be mentioned.
AC30: In the revised acknowledgment, both Ingjerd Haddeland and Bernhard Lehner have been acknowledged for providing data.

Technical corrections
All of approx. 35 “technical corrections” suggested were done except the following:
Abstract not divided into 2 paragraphs as this is unusual for abstracts.
With respect to Table 2, only parts of the recommendations have been followed. No equations are added as that would require an additional table or a lot of space in the text.
No new paragraph in (old) page 4790, line 24, would only have two sentences.
Fig. 4: colours not changed.
Fig. 9: A note was added to figure caption that legend for Volga also applies to Missouri and Colorado.

Referee 3

Specific comments
RC31: P4781 L17 “subtracting total consumptive water use from water stored in lakes. . .” How did you set the volumetric capacity of lakes? The whole water body of lakes can be used as water resources? Are lakes depleted when overexploited?
AC31: Not the whole natural lake can be used a water resources, only within a storage variation of 5 m. In WGHM, lake levels can generally vary by 5 m. If lakes drop below a certain level (0), no more water is withdrawn.

RC32: P4781 L23 "Global consumptive water use has more than doubled between 1951 and 2002..." You described that you used domestic, industrial and livestock water use of the year 2002 and 1961-1990 climatic data to estimate irrigation water use. How did you estimate water use between 1951 and 2002?
AC32: Time series of consumptive use between 1951 and 2002 is used for model tuning. It is estimated in the different water use modules of WaterGAP. In the case of domestic water use, for example, the time series are derived from time series of population and per-capita water use (which for some point in time is taken from statistical data), the per-capita water use being a function of structural (GDP) and technological change. Irrigation water use is a function of climate as well as of irrigated area, and we have estimated time series of irrigated area per country, mainly based on FAO data.

RC33: P4782 L9 "If on any day there is not enough water available in surface waters to satisfy the consumptive use, the model will take out this consumptive water use later in the year or in the next year. This approximates withdrawals from renewable groundwater resources. . ." First, this assumption seems to carry over not only groundwater but also river discharge. Why it approximates renewable groundwater? Second, this is a big assumption because water deficit in dry period is canceled out by surface water in wet period even if there is no reservoir. In other words, this assumption acts as virtual reservoir, dumping temporal variation of surface water and water demand. If you agree this, you need to mention this in text. Here a question came to my mind: What happens if you disabled this carry over assumption?
AC33: To address you questions and concerns, the following sentences have been added to the revised manuscript:

“This approximates withdrawals from renewable groundwater resources, as, in reality, groundwater can be withdrawn even in periods with low river flows. The delayed satisfaction of water requirements leads to a stronger discharge reduction in the river (as compared to simulations where delayed satisfaction is not allowed), which is more realistic if groundwater that is connected to surface waters is actually withdrawn. Also, groundwater withdrawals in a certain month lead to delayed response of baseflow to the river, which is also somehow reflected by the delayed satisfaction approach. Nevertheless, actual temporal variations of the effect of water withdrawals on surface water flows are only approximated very roughly.”

RC34: P4783 L1 “The new reservoirs and regulated lakes data set was derived by adding additional reservoirs from a preliminary version of the GRanD database. . .” GRanD database needs to be further explained. You mentioned that GRanD database provides 6568 reservoirs and 52 regulated lakes, but the number of reservoirs is far below the total number of dams and reservoirs in the world (45000-60000). Which types of
reservoirs does GRanD database cover? Are there any geographical biases?
AC34: see AC35

RC35: P4783 L22 "The reservoir operation algorithm of Hanasaki et al. (2006) was implemented in WGHM for the 1074 global reservoirs and regulated lakes" How did you select 1074 out of 6568+52 reservoirs and lakes?
AC35: In the revised version, par. 2 of Section 2.1.3 reads:
“The new reservoirs and regulated lakes data set includes 6568 reservoirs and 52 regulated lakes (Table 1 and Fig. 2). It was derived by adding additional reservoirs from a preliminary (July 2008) version of the GRanD database (Lehner et al., 2008) to the 886 reservoirs included in WGHM 2.1f. For WHGM 2.1f, the Global Lakes and Wetlands Database of Lehner and Döll (2004) had already been augmented by 64 reservoirs. The development of the GRanD data set aimed at including, as polygons, all reservoirs world-wide with a storage capacity of more than 0.1 km3. While for Europe and the USA, this goal was probably reached, it is believed that the dataset is incomplete particularly in China, India and South America. GRanD does not distinguish between regulated lakes and reservoirs. Therefore, all “global” reservoirs (with an area of more than 100 km2 or a maximum storage volume of at least 0.5 km3) were checked to decide whether they are actually regulated lakes. This resulted in the identification of 52 regulated lakes, in addition to 1022 “global” reservoirs. “Global” reservoirs and lakes are assumed to be fed by river discharge from the upstream cell, while smaller reservoirs are assumed to be “local”, i.e. they are only fed by the runoff generated within the grid cell.

RC36: P4783 L26 "Different from Hanasaki et al. (2006) ..." Until I read Hanasaki et al. (2006) very carefully, I couldn’t understand what is different at all. You need to add a brief description of the algorithm of Hanasaki et al. (2006) and list up what points are different from their original algorithm.
AC36: The reservoir algorithm of Hanasaki et al. has been explained in the revised version in more detail, and the differences to the WGHM reservoir algorithm are listed more extensively and clearly. The revised fourth paragraph of section 2.1.3 now reads:
“Hanasaki et al. (2006) developed two different algorithms, one for reservoirs with irrigation as their main purpose, and another for all other reservoir types. In both cases, annual release is a function of the long-term average annual reservoir inflow and relative water storage at the beginning of the operational year. The operational year has is computed based on the seasonal flow dynamics since no data are available. Different from Hanasaki et al. (2006), in WGHM the releases are a function of the long-term average value of reservoirs inflows plus the difference between precipitation and evaporation over the reservoir, as the long-term average annual outflow of a reservoir depends not only on the inflows but also on the reservoir water balance, in particular in case of large reservoirs (and regulated lakes). In the case of non-irrigation reservoirs, monthly outflows are assumed to be constant throughout the operational year. The monthly fluctuation of releases of irrigation reservoirs

depends on monthly downstream consumptive water use, taking into account water use in the next five downstream cells, or down to the next reservoir or the river mouth. This is different from Hanasaki et al. (2006), who took into account water use in a maximum of 10 1° grid cells downstream, and preferred water withdrawals to consumptive use. With the shorter range in the WGHM version, we mostly avoid that there is more than one reservoir that could provide water for a certain cell. Overflow occurs if reservoir storage capacity would be exceeded. Consumptive water use is only subtracted from reservoirs if reservoir storage exceeds 10% of storage capacity."

RC37: P4785 L25 "Richter et al. (1997) require daily discharges, . . . Therefore, we only considered indicators that are based on monthly and annual river discharge estimates" What is the spatiotemporal scale of the original works of Richter et al. (1997) and Black et al. (2005)? Because the discussion of aquatic ecosystem is scale dependent, you need to explain the logic here more carefully.
AC37: The following sentences were added to section 2.2: "The Indicators of Hydrologic Alteration set of Richter et al. (1997) was developed to guide the operation of individual reservoirs. In these cases, daily discharge measurements are generally available. Therefore, most of the 32 indicators proposed by Richter and colleagues require daily discharges and thus cannot be computed well by a global hydrological model that is driven by monthly climate input data."

RC38: P4794 L23 "while irrigation has lead to a decrease of only 1.5% as compared to only rainfed agriculture" I'm wondering this is attributable to the carry over assumption of water deficit. The model of Rost et al. (2008) didn't adopt this assumption, while your model did. Usually the amount of water withdrawal from streamflow adopting the former modeling falls below that of the latter.
AC38: Yes, this should be at least one of the reasons. The following sentences have been added: "The discrepancy between the two model results may be due to the fact that in the model of Rost et al. (2008), delayed satisfaction of water requirements from surface waters, which approximates water withdrawals from renewable groundwater resources, is not implemented, such that a smaller fraction of the water requirements can be fulfilled than in WGHM (see Fig. 1b)."

Technical corrections
P4812 Figure 9. Some of the legends of lines are missing in the figure of the Missouri and Colorado River.
Fig. 9: A note was added to figure caption that legend for Volga also applies to Missouri and Colorado.